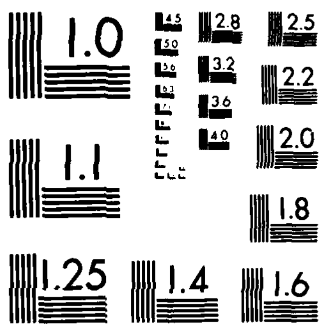
AD-A151 977 A COMPARISON OF ALTERNATIVE ANALYTIC MODELS FOR EVENT RELATED POTENTIAL RECORDS(U) WASHINGTON UNIV ST LOUIS MO DEPT OF PSYCHOLOGY E HUNT ET AL. NOV 84 1/1

UNCLASSIFIED AFOSR-TR-85-0249 AFOSR-83-0289 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AFOSR-TR- 85-0249

4

AD-A151 977

A COMPARISON OF ALTERNATIVE ANALYTIC MODELS FOR EVENT RELATED POTENTIAL RECORDS

FINAL REPORT

EARL HUNT AND PRAPAN TIANWATTANATATA

NOVEMBER 1984

This research was sponsored by:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Project-Task 2313/A4

Grant No. AFOSR-83-0289

85 03 12 109

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 85-0249	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A COMPARISON OF ALTERNATIVE ANALYTIC MODELS FOR EVENT RELATED POTENTIAL RECORDS		5. TYPE OF REPORT & PERIOD COVERED FINAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) EARL HUNT and PRAPAN TIANWATTANATATA		8. CONTRACT OR GRANT NUMBER(s) AFOSR-83-0289
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Washington Seattle, Washington 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2313/A4 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NL Bolling AFB, DC 20332		12. REPORT DATE November 1984 13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Principal component analysis, PCA, confirmatory factor analysis, factor analysis, event related potential, evoked potential, time series analysis, psychophysiology.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Principal Component Analysis is a technique that is widely used to extract component wave forms from event related potential (ERP) records. Analysis of simulated ERP records indicates that Principal Component Analysis may produce biased solutions in some cases. Two alternative methods of analysis are considered; confirmatory factor analysis and time series analysis. Confirmatory factor analysis provides superior results if the experimenter has reason to reject some component wave forms on a priori grounds. Time

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

series analysis is preferable in situations in which the analysis can be conducted on only a few records.

Final Technical Report
AFOSR-83-0289
November 1984

A COMPARISON OF ALTERNATIVE ANALYTIC MODELS FOR EVENT RELATED POTENTIAL RECORDS

Department of Psychology
University of Washington
Seattle, WA 98195

Dr. Earl Hunt

Controlling Office: Air Force Office of Scientific Research/NL
Bolling Air Force Base, DC 20332

A Comparison of Alternative Analytic Models for Event Related Potential Records

Earl Hunt and Prapan Tianwattanatata

The University of Washington

The Event Related Potential (ERP) is a record of the electrical activity detected in the brain following the presentation of a stimulus. The ERP is assumed to be generated by the summation of a number of component wave forms. Some of these wave forms reflect the brain's response to the physical characteristics of the stimulus. Other, late arising components are believed to be associated with the brain's interpretation of the stimulus (Donchin, 1975). Therefore the form of the later components of the ERP, and changes in those forms as a result of experimental manipulation, are of considerable interest.

In humans the ERP is normally recorded from scalp electrodes, and contains a considerable amount of electrical noise in addition to the signal generated by the components wave forms. Let x[i](t) be the observed potential in ERP record i time t, where t is measured from the presentation of the stimulus. The mathematical model *usually assumed for the ERP is*

$$x[i](t) = \sum_{k=1}^{K} g[i,k](t) + e[i](t) \quad (1)$$

where g[i,k](t) is the value of the kth of K underlying component wave forms in record i at time t, and e[i](t) is an error term for record i at time t. The goal of analysis of ERP records is to identify the underlying components, i.e. the form of the functions {g[i,k](t)}.

In order to do this several simplifying mathematical assumptions are made. Time is quantized into T discrete intervals, so that (1) may be regarded as an expression of a relation between elements of T dimensional vectors (T > K) rather than as an expression of a relation betwen continuous functions. The component wave forms are assumed to be non-zero only over part of the period of recording. Mathematically, for each component wave form it is assumed that there are characteristic time periods, T'[k] and T"[k] *such that*

$$g[i,k](t) \; \underline{<>} \; 0 \text{ if and only if } T'[k] > t > T''[k]. \quad (2)$$

Less mathematically, each component wave form is assumed

to begin and end at some time during the recording epoch.

In many analyses it is assumed that the N recordings (i = 1..N) can be regarded as replications of the same experiment, i.e. that they differ only in the noise components {e[i](t)}. If this assumption is accepted each component wave form can be thought of as consisting of two subparts, a standard form that varies in magnitude from plus one to minus one, and that is constant across records, and an amplitude term that is characteristic of the combination of component and record. Algebraically,

(3) $g[i,k](t) = A[i,k]\ y[k](t).$

The coefficients {y[k](t)} establish the form of the component wave across all records, and the term A[i,k] *establishes the size of the kth component form in the ith* record. Substituting (3) into (1), the equation for the ERP record becomes

$$(4) \qquad x[i](t) = \sum_{k=1}^{K} A[i,k]\ y[k](t) + e[i](t).$$

Given knowledge of the N vectors of observed values, **x** [i], (i = 1..N), equation (4) can be solved for the

values of the A and Y terms by a technique known as Principal Component Analysis (PCA) defining the component wave forms, (Donchin and Heffely, 1979). The mathematical technique itself is well known (Mulaik, 1972), and will not be further described. By using PCA, one implicitly makes more assumptions about data generation than are implied by the reasoning that lead to (4). These are:

(1) The amplitudes of any two components, { A[i,k] } and {A[i,k']}, are uncorrelated across records.

(2) The error terms { e[i](t) } are uncorrelated across both records and time points.

(3) The same standard wave form, i.e. the same coefficients {y[k](t)} apply across all records.

(4) Both the amplitude and error terms are multivariate normally distributed.

Hunt (in press, see also Hunt, 1984) has discussed *the reasonableness of these assumptions as statements* about the way in which ERP records are generated. Assumptions (1) and (2), in particular, were questioned. Since two ERP components, within a record, reflect electrical activity within the same brain, any process

that leads that brain to have a characteristic amplitude of electrical activity across all processes will produce correlatons between amplitudes. Since error terms are produced by extraneous events occurring at a particular time, their effects may last, with diminishing magnitude, for several time periods. This would induce a correlation betwen error terms across time periods. Similar concerns about the reasonableness of the PCA assumptions as applied to ERP analysis were raised by Wood and McCartny (1984).

Hunt explored the seriousness of violations of assumptions (1) and (2) by using PCA to analyze computer generated simulated ERP records with known component wave forms and error terms. The simulated records themselves were derived from similar simulation analyses conducted by Wood and McCartny. Records were generated that either conformed to the PCA analysis, conformed except for a violation of assumption (1), or which violated both assumptions (1) and (2). Violations of assumptions (3) and (4) were not explored. Three criteria were used to evaluate the results. The recovered component wave forms (the 'y' terms) were compared with the wave forms used to generate the data. PCA results were surprisingly robust. Little difference was observed between the original and the derived wave forms. The percentage of variance in the ERP records, over time, accounted for by each wave form,

was compared in the original records and the records derived from PCA. This can be looked upon as a measure of the relative sizes of the component wave forms, averaged across records. The agreement was again surprisingly good, even for data that was generated by methods that violated both assumptions (1) and (2). The amplitudes of component wave forms used to generate records were correlated with the amplitudes assigned to those records by PCA. Here violations of assumptions did matter. The correlations were high when the PCA assumptions were met, but dropped sharply when they were not. This result poses a serious problem, because it would influence the validity of an anlysis of variance of PCA component amplitudes obtained under different conditions.

In this paper we consider two alternatives to the conventional PCA method for identifying component wave forms. This first method is a statistical procedure called Confirmatory Factor Analysis (COFAM). The second method is a technique taken from econometrics, called Time Series Analysis (TSA). COFAM can be thought of as an alternative to PCA. TSA is a procedure for preparing data prior to applying PCA. The alternative methods will be evaluated by analyzing the simulated data that was used in Hunt's study. (We note again that this data is closely related to that used by Wood and McCartny.) The results of the COFAM

and TSA analyses of the data will be compared to the results produced by PCA alone.

THE SIMULATED DATA

Equation (4) was used to generate simulated ERPs based on the wave forms used by Wood and McCartny, with a minor exception. Wood and McCarthy generated records over 64 time points. The records used here contained 20 time points, *equally spaced over the range* used by Wood and McCarthy. The limitation to 20 points was forced on us by limitations in the COFAM and TSA computer programs used. It should not influence the results, since 20 measures are more than sufficient to define three components. Figure 1 shows the resulting wave forms. Reading from the left to right, Component I is a symmetrical component that begins early in the record, and rises and falls about its maximum amplitude. Component II is a similar symmetrical component wave that begins immediately after component I ends. Component III is a continuously rising wave that begins at the same time as does component II, and continuously ascends in a negatively accelerated manner throughout the *record*.

In generating their data Wood and McCarthy used a rather small error variance, so that the component wave

forms accounted for 98% of the variance in the data. The simulations reported here used an error variance equivalent to approximately 25% of the total variance in the data, i.e. a signal to noise ratio of three to one.

Figure 1

Three data sets were generated. Data Set I was generated as described above. Data Set II was generated by a modification of Equation (4) that produced correlated component amplitudes. A general term A[i,0] was chosen at random for each record, and was added into the amplitude term used to establish the individual data points. The data generation equation becomes

$$(5) \quad x[i](t) = \sum_{k=1}^{K} (A[i,k] + A[i,0)y[k](t) + e[i](t).$$

The value of A[i,0] was chosen randomly from a normal distribution whose variance was manipulated to produce a correlation of .4 between component amplitudes, calculated across records. Thus Data Set II produced a violation of statement S.1, the orthogonality assumption of PCA.

Data Set III was generated using both correlated components and correlated errors (i.e. violation of S.4).

Each record in Data Set II can be thought of as the sum of two vectors. The first vector, whose elements are established by the summation term in (5), is determined by the component wave forms. The second is the residual or "error" vector, established by the terms { e[i](t) }, which are chosen independently over i and t. In Data Set III revised residual vectors were derived from the original ones by the following rules:

(6a) $e[i]'(1) = e[i](1)$.

(6b) $e[i]'(2) = .5\ e[i](2) + .25\ e[i](1)$.

(6c) for $i > 2$,

$$e[i]'(t) = .5\ e[i](t) + .25\ e[i](t-1) + .125\ e[i](t-2).$$

The revised residual terms are correlated across time (t), because the term at each time point is a stochastic function of the terms introduced at previous time points. The rules of equations (6 a-c) were chosen to mimic a situation in which the effect of an extraneous ("error") event on the ERP recorded extends, with diminishing intensity, over several time periods.

A SUMMARY OF PREVIOUS RESULTS

Since the purpose of this paper is to compare the results of PCA analyses of the three data sets to the analyses obtained with COFAM and TSA, a brief summary of the results reported in Hunt (1984) is in order. These results serve as a standard of comparison for the results reported in the following sections.

PCA followed by varimaxing was applied to each of the data sets. Three results will be reported; the component wave forms recovered from the data, the percentage of variance in the wave form that was assigned to each component, and the correlation between the component amplitudes used to create the simulated records and the amplitudes recovered by PCA. The correlation was calculated across records. These are the statistics reported by Wood and McCartny.

Figure 2,3,4

Figure 2 shows the wave form recovered by the PCA for Data Set I, in which the data was generated by the PCA model. The component wave forms agree well with the ones shown in Figure 1. Figures 3 and 4 show the wave forms constructed from Data Sets II and III. The results are similar for each case. The shape of component wave form I, the *non-overlapping component*, is recovered much as before. By contrast, the two overlapping wave forms,

components II and III, are not reconstructed as well. Component II is changed from a symmetrically rising and falling wave form to one with a rapid rise and a slow decline to an intermediate level. Component III begins its rise later than it should, and does not reach its full height.

Table 1 shows the percent variance in the wave forms that is allocated to variance in each of the three components. The first line of the table presents the figure for the original data, the second, third, and fourth present the variance allocated to the components recovered by the three applications of PCA. The analysis recovered the relative sizes of the different components rather well. The largest 'miss' was a 9% misallocation of variance away from a component III for Data Set I.

Tables 1 and 2

Table 2 presents the correlations between the component amplitudes used to generate each record and the time amplitudes recovered by PCA. In the case of Data Set I the recovery is quite accurate. The drop in correlation below 1.0 is about what would be expected given the amount of error variance introduced in the simulation. The results from Data Sets II and III are more problematical. Two statistics are presented for

each component. The first is the correlation between the PCA estimates of component amplitudes and the component amplitude used to generate the record, *including the general component*. *In terms of equation* (5), this figure is the correlation between the PCA estimate and the term (A[i,0] + A[i,k]) for k = 1,2,3. The term in parentheses in Table 2 is the correlation between the PCA estimate and the component amplitude excluding the general component, i.e. the term A[i,k]. As can be seen, the second correlation is substantially below the first. What this means, in practice, is that if the ERP is generated by several component wave forms, and if two or more of these are susceptible to influences from the same variables (thus inducing a correlation between component amplitudes, across records), then PCA estimates of component amplitudes may be of reduced accuracy.

THE APPLICATION OF CONFIRMATORY FACTOR ANALYSIS METHODS

Confirmatory factor analysis is a member of the general class of techniques known as "factor analysis." PCA is included in this family (Mulaik, 1972). The mathematical basis of COFAM is considerably more involved than that for PCA. Full descriptions have been provided by Joreskog and Sorboom (1979) and by Long

(1983). Here an intuitive contrast between PCA and COFAM will be offered.

The mathematical problem of identifying component wave forms from data generated by equation (4) has no unique solution. In PCA a unique solution for the problem is defined by making the assumption that there is no correlation between component amplitudes, and by applying a mathematical technique known as varimaxing following an identification of the number of components by direct application of PCA. These will be referred to as the <u>orthogonality</u> and <u>varimaxing</u> restrictions. See Hunt (1984) or any of the other previously presented references for details. The analysis finds values for the component wave forms *across all component and time periods* (i.e. the set { y[k](t) }).

In COFAM both the orthogonality and varimaxing *restrictons are removed. However, they must be replaced* by other restrictions in order to define a mathematically solvable problem. The general COMFAM method allows one to specify a variety of restrictions. *In the ERP* application discussed here the restriction used was that the non-zero coefficients of the various component wave forms were restricted to a particular

time interval. In physiological terms, this amounts to an assumption that certain components do not appear outside of the specified time interval. For example, suppose that an investigator were to be interested in the P300, a late appearing component wave form that has been associated with a variety of psychological phenomenon (Donchin, 1975). In a conventional PCA the mathematical technique permits the appearance of the P300 at any time during recording. Based on psychophysiological evidence, though, an investigator might reject the idea that the P300 component could appear until at least 100 milliseconds after stimulus presentation. The COFAM method allows the investigator to build in scientific knowledge about the P300, by instructing the computing method to consider only those solutions that fit the indicated restrictions on wave forms.

In the simulations considered here the wave forms *sought were restricted* to the "correct" range, i.e. non-zero amplitudes for each component form were permitted only for time intervals where non-zero amplitudes had, in fact, been used to construct the data. This ensures that *the answer obtained by COFAM* will be the best possible answer. Less accurate answers would be produced by less accurate assumptions. (In

fact, if no assumptions about intervals are made, and the orthogonality assumption is made, then COFAM can be used to produce a model that very closely resembles the PCA model.) In a way, the simulations can be looked upon as those that would be produced by an extremely confident researcher whose assumptions were correct! We have chosen this unrealistically omniscient approach to illustrate the best possible COFAM analysis. In practice, a researcher would leave a margin of error, by permitting non-zero coefficients over all intervals in which a predicted component might be found, even if the researcher expected the component to actually be located in a smaller interval. If this were the case COFAM would produce three classes of component coefficients; values that were clearly not zero over the true range of the component, values that were "statistically close" to zero over the range of time points that were outside the true interval of the component but inside the interval considered by the researcher, and dictated zero coefficients outside the researcher's permitted interval.

COFAM was applied to Data Sets *I*, *II*, and *III* in the manner just indicated. Figures 5, 6, and 7 show the wave forms that were derived. They are clearly much closer in form to Figure 1 than are the PCA derived

wave forms shown in Figures 2, 3, and 4.

Table 3 shows the percentages of variance in the ERP records assigned to each component wave form in each data set. These should be compared to the "right answer" shown in line 1 of Table 1. These estimates are somewhat more variable than the estimates obtained by PCA. The largest error is a 15.5% underassignment of variance to Component II in Data Set II. The chief problem in misassignment seems to be due to misassignments of variance to components that overlap in time, a phenomenon also noted by Wood and McCarthy.

While we do not know exactly why the PCA method is somewhat more accurate than the COFAM method, the result is not exceptionally surprising. COFAM is based upon the maximum likelihood method of parameter estimation, which is somewhat less accurate than the minimization of least squares method that is used to estimate parameters in PCA.

Table 4 shows the correlations between the component amplitudes used to create the records and the component amplitudes estimated by PCA. Two correlations for each estimate were calculated for data sets II and

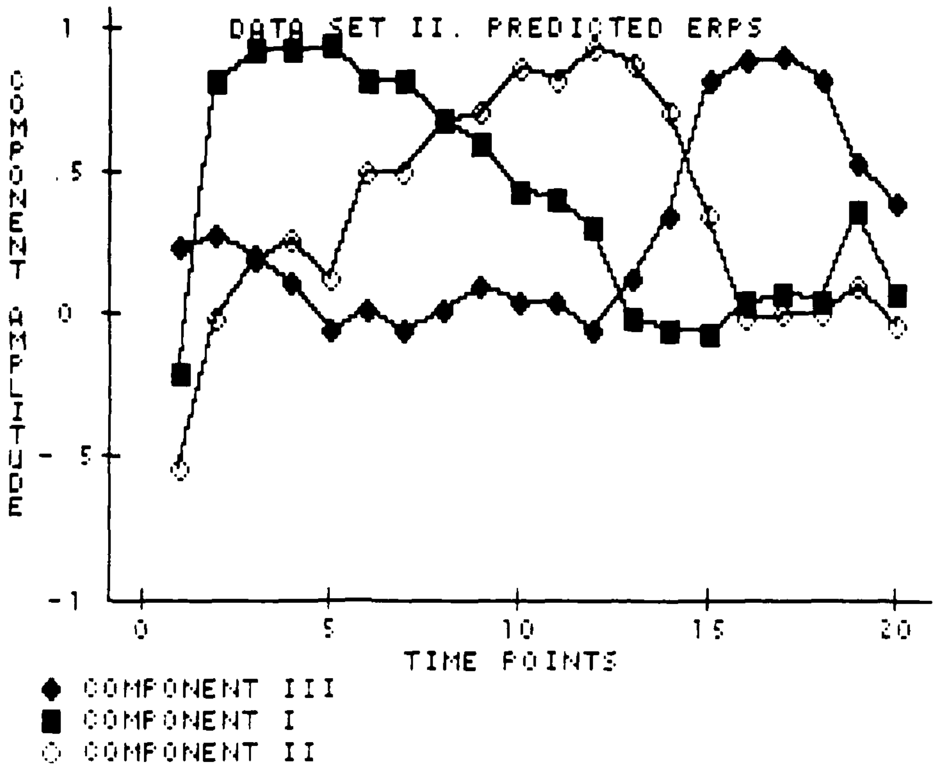


Figure 9B

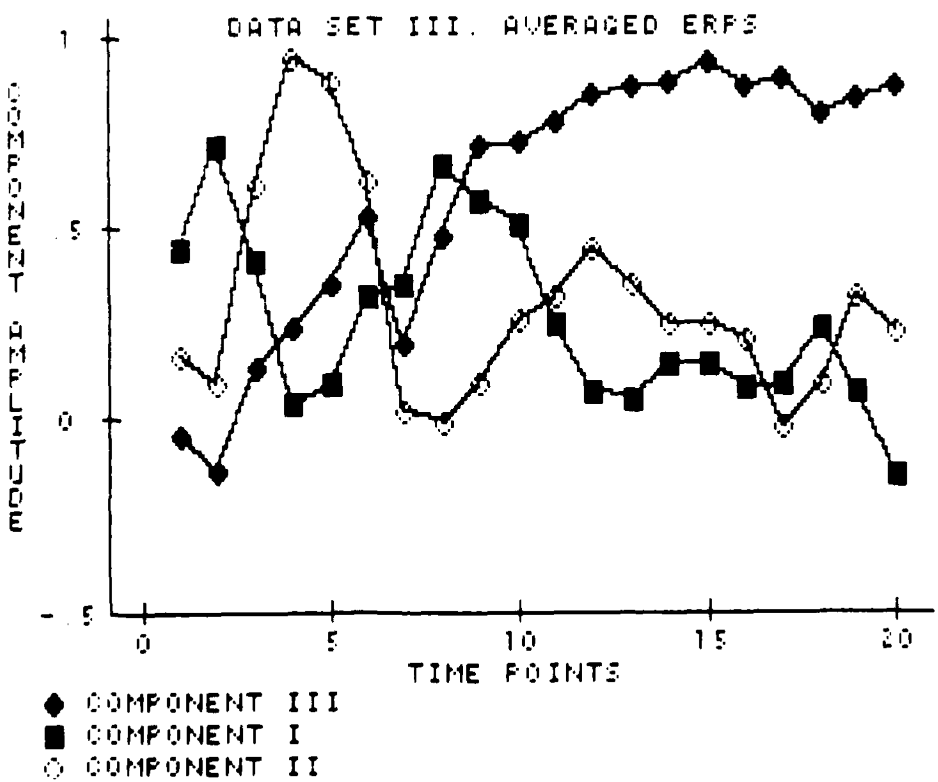


Figure 10A

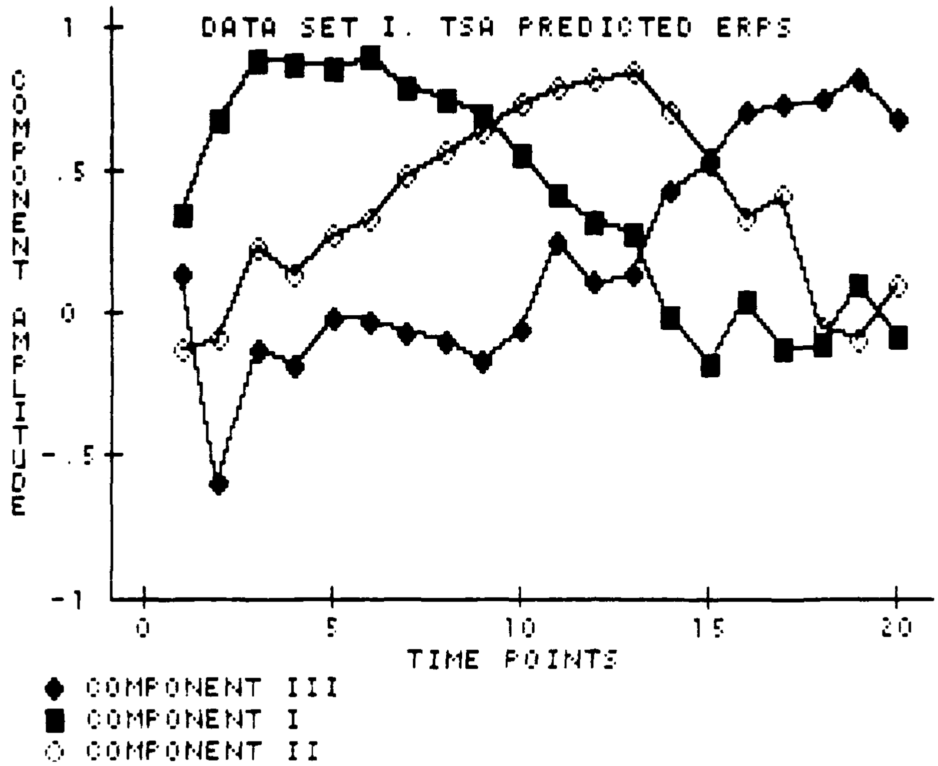


Figure 8B

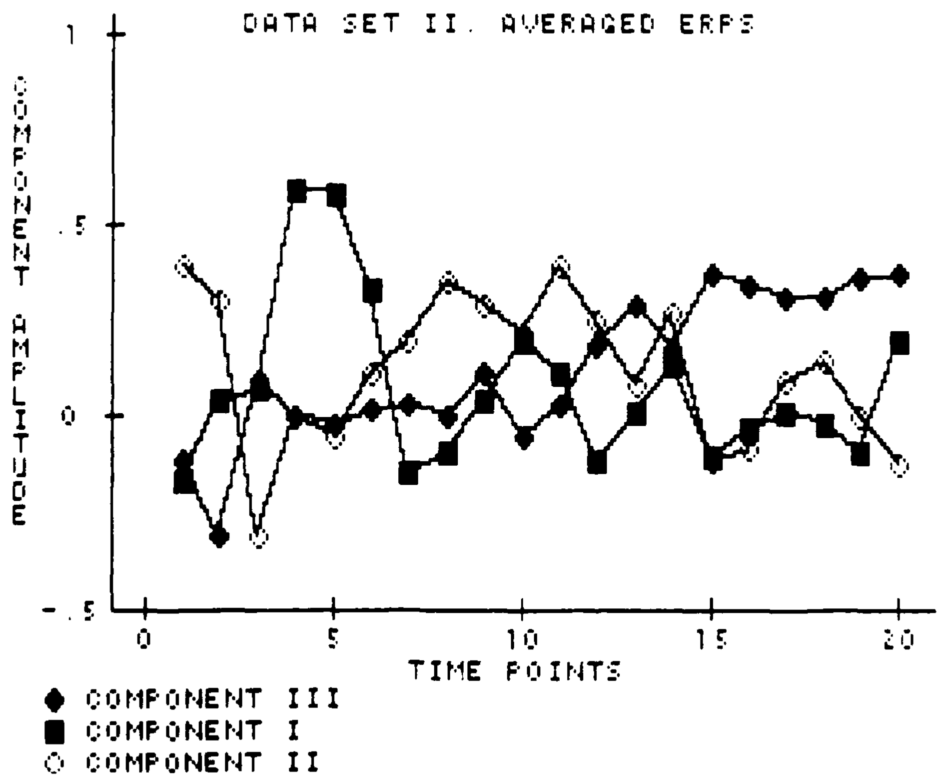


Figure 9A

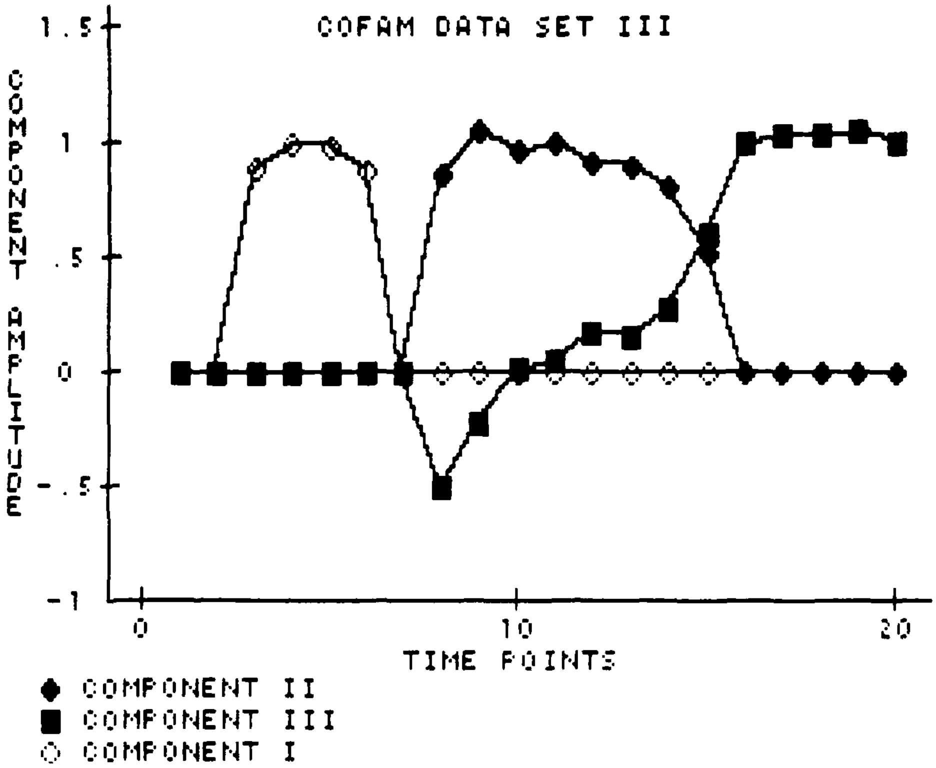


Figure 7

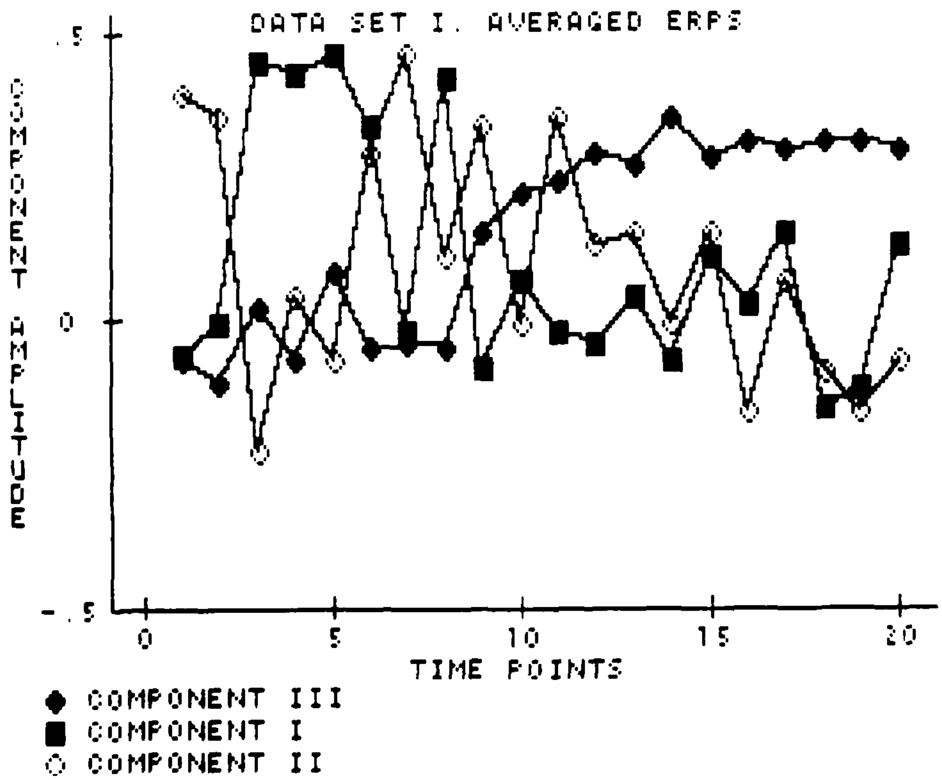


Figure 8A

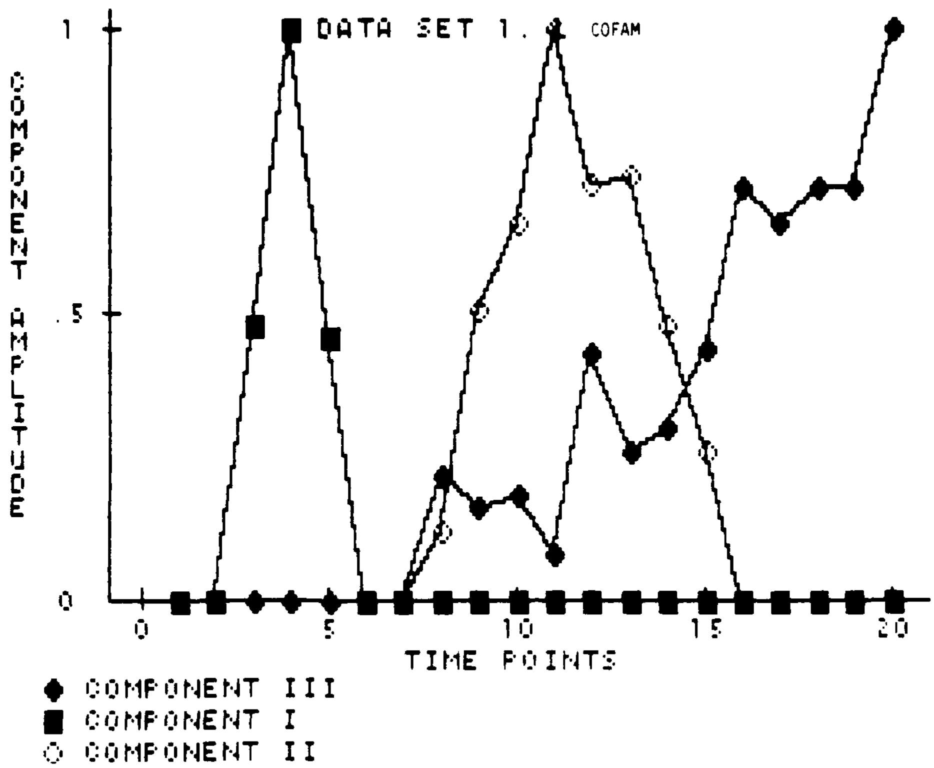


Figure 5

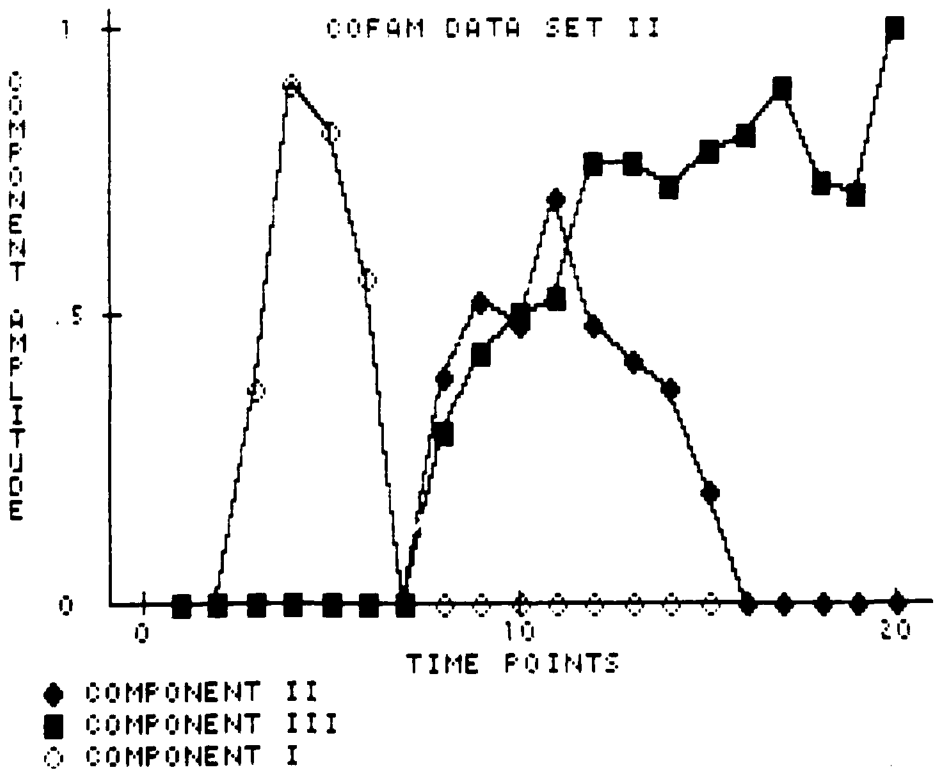


Figure 6

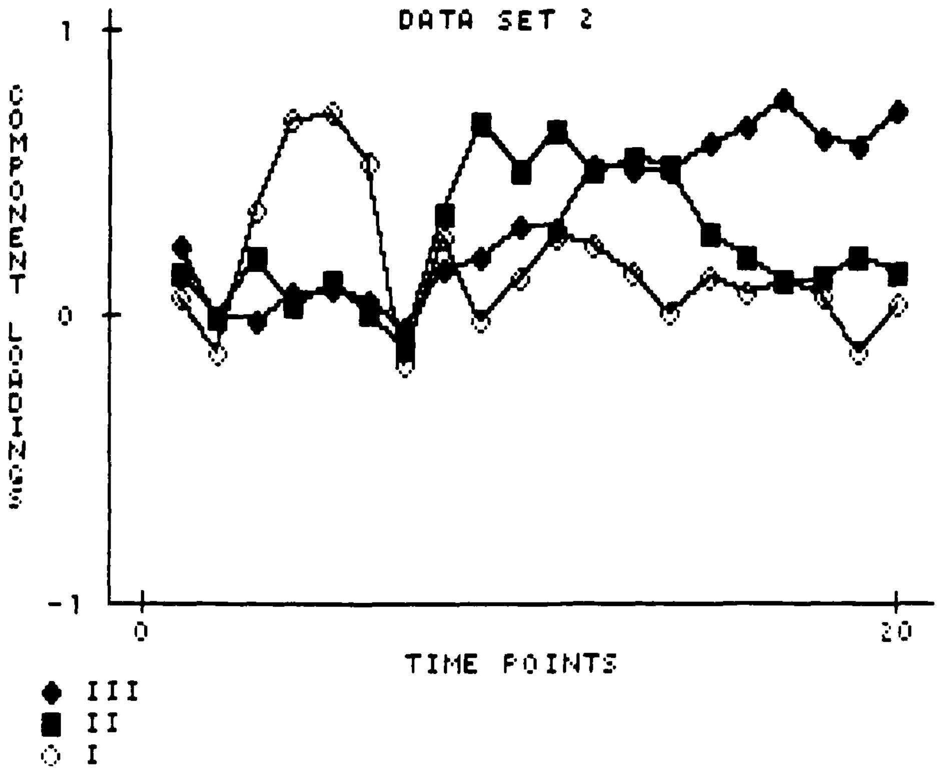


Figure 3

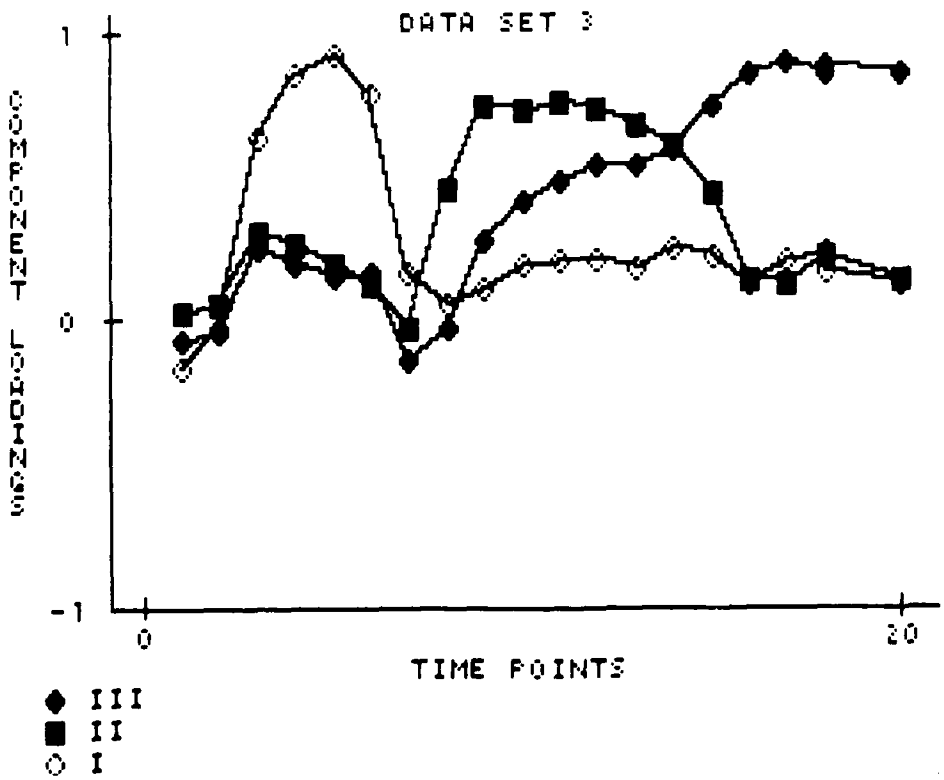


Figure 4

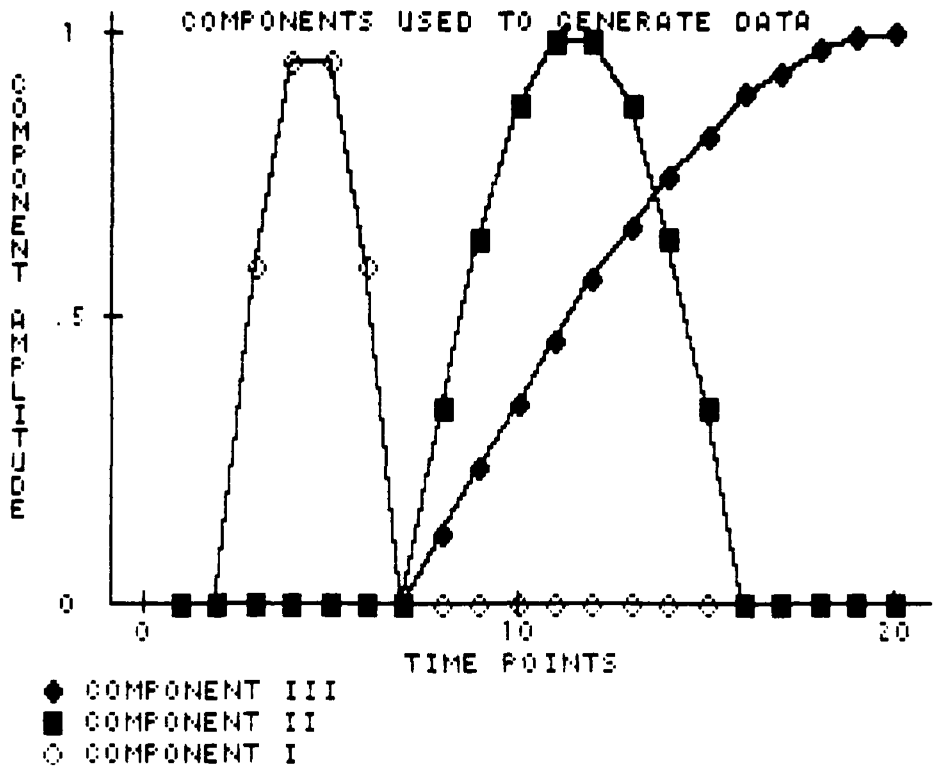


Figure 1

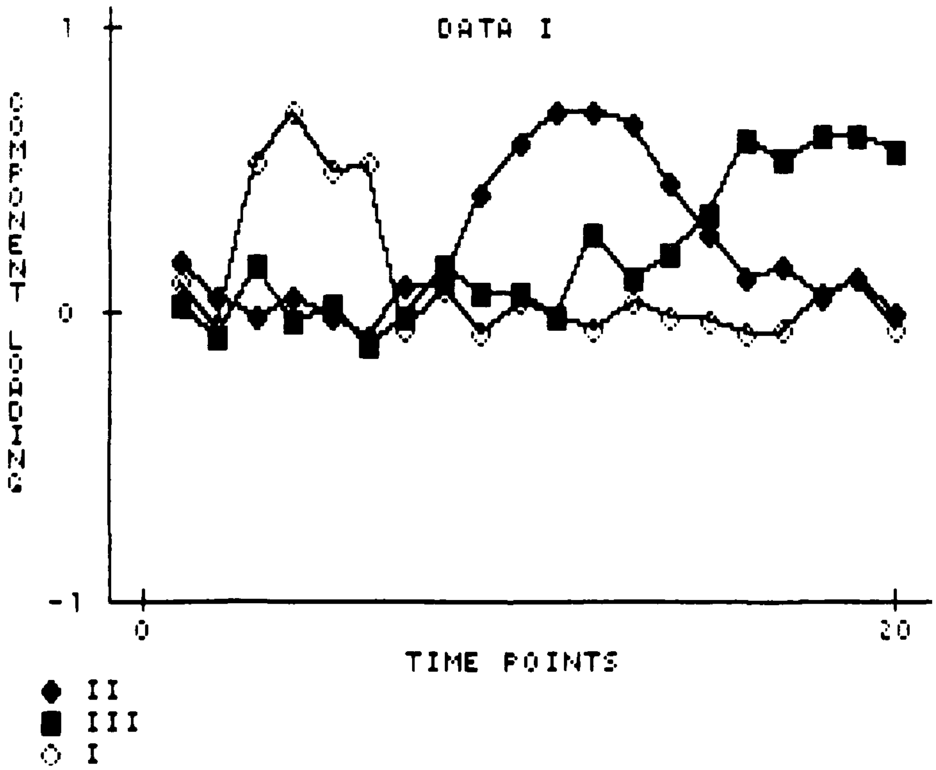


Figure 2

Figure Legends

1. The three component wave forms used to construct the simulated data. Wave forms are referred to as Forms I, II, III, from left to right.

2. The component wave forms recovered by PCA from Data Set I.

3. The component wave forms recovered by PCA from Data Set II.

4. The component wave forms recovered by PCA from Data Set III.

5. The component wave forms recovered by COFAM from Data Set I.

6. The component wave forms recovered by COFAM from Data Set II.

7. The component wave forms recovered by COFAM from Data Set III.

8. The component wave forms recovered by applying PCA to idealized wave forms. Data Set I.

9. The component wave forms recovered by applying PCA to idealized wave forms.
Data Set II.

10. The component wave forms recovered by applying PCA to idealized wave forms. Data Set III.

Table 2

Condition	Component I	II	III
Data Set I	.82	.79	.83
Data Set II	.82 (.65)	.82 (.64)	.90 (.71)
Data Set III	.83 (.71)	.85 (.60)	.95 (.78)

Correlations between the amplitudes of components used to generate each record and the component amplitude assigned to that record by PCA in each of the data sets.

Table 3

Condition	Component I	II	III
Recovered, Set I	21.1	39.3	39.6
Recovered, Set II	18.9	15.6	65.5
Recovered, Set III	26.1	20.7	53.2

Percent variance of wave forms assigned to each component in the model used to generate the data and recovered by COFAM after analysis of each of the data sets.

Table 4

Condition	Component I	II	III
Data Set I	.76	.79	.84
Data Set II	.65 (.85)	.60 (.80)	.31 (.82)
Data Set III	.60 (.92)	.31 (.82)	.72 (.96)

Correlations between the amplitudes of components used to generate each record and the component amplitude assigned to that record by COFAM in each of the data sets.

Wood, C. C. and McCarthy, G. (1984). Principal component analysis of event-related potentials: Simulation studies demonstrate misallocation of variance across components. Electroencephalography and clinical neurophysiology. 59, (3), 249-260.

Table 1

Condition	Component I	II	III
Assigned by Model	17.9	32.1	50.0
Recovered, Set I	23.6	35.7	40.8
Recovered, Set II	21.9	29.5	48.5
Recovered, Set III	23.9	29.9	46.1

Percent variance of wave forms assigned to each component in the model used to generate the data and recovered by PCA after analysis of each of the data sets.

does the conventional averaging procedure.

Figures 8, 9, 10 here

CONCLUSION

The simulation studies of Hunt and Wood and McCarthy indicated that the conventional PCA procedure could introduce serious biases into the identification of various statistics defining component wave forms. The results presented here indicate that alternative analyses may produce considerably more accurate results. In situations in which an investigator has strong grounds for believing that only certain types of component wave forms are possible, serious consideration should be given to using this knowledge, via the COFAM technique, as an aid in defining the actual componnt wave forms. In situations in which averaging over large number of trials is suspect, the TSA method of producing idealized records as input to PCA appears to be superior to a simple averaging procedure.

References

Box, G. E. P. and G. M. Jenkins (1976) Time Series Analysis: Forecasting and Control : San Francisco: Holden-Day.

Donchin, E. (1975) Brain electrical correlates of pattern recognition. In G. F. Inbar (ed.) Signal analysis and pattern detection New York: Wiley

Donchin, E. and Heffley, E. F. (1979). Multivariate analysis of event related potential data: a tutorial review. In D. Otto (ed.) Multidisciplinary perspectives on event related potential research. EPA-600/9-77-043: Washington, D.C. U.S. Government Printing Office.

Hunt, E. (1984) Mathematical models of the event related potential. U. Washington Psychology Department technical report. Project-Task 2313/A4, Grant AFOSR-83-0289

Hunt, E. (in press) Mathematical models of the event related potential. Psychophysiology

Suppose that one is given observations for N ERP records, each consisting of R cycles of length T, i.e. R x T observations per record. Equation (7) can be used to construct a predicted "R + 1st" cycle that represents the cyclic information that has been cumulated over first R actual repetitions. In the ERP analysis, one might regard the R+1st cycle as an ideally stable cycle. It is analagous to an econometrician's predictions for the next fiscal year, based upon data from the last R years. Subsequently, when we refer to a "TSA predicted record" we shall mean a vector of T simulated ERP observations produced in this manner.

The goal of this study was to compare the accuracy of PCA derived wave forms based upon TSA predicted records to the accuracy of the same analysis conducted on records derived from the more conventional averaging procedure. A simulation very similar to the previous simulation was conducted. Five hundred simulated ERP records were constructed, each consisting of 20 observations. The records were developed once using the techniques for Data set I, once using the Data Set II techniques, and once using the Data set III techniques. The five hundred records were randomly partitioned into

twenty cases of twenty five records each. Thus a case is analagous to a subject, and a record is analagous to the recordings from a single trial. For each case, two ideal records were constructed; once by simply averaging across records and once by the TSA method just described. Component wave forms for the twenty ideal records were then computed using the PCA plus varimaxing procedure.

Figure 8A shows the component wave forms deived from the analysis of averaged records from data set I. The desired components are present, but the pattern is unclear, to say the least. This is probably entirely due to the reduced effective N, 20, compared to the effective N of 100 that was used to *develop* Figure 2, the conventional PCA procedure applied to this data. Figure 8B shows the component wave forms derived from the same data, but applying TSA instead of averaging to produce the ideal records. Although there is a good deal of noise in the data, Figure 8B is clearly a closr approximation to Figure 1 than x Figure 8A.

Figures 9 and 10 present the same comparison for data sets II and III. In each case preparation of the ideal records by TSA provides a more accurate construction of the original component wave forms than

subsequent T time intervals. (To complete the analogy to economics, think of an economic indicator that is recorded monthly, with the "stimulus presentation" being the start of a fiscal year.) The following assumptions are made:

1. A regular series of brain events is initiated following the presentation of each stimulus. The ERP record thus contains a component that is cyclic, begins with the stimulus presentation, and has period T.

2. At each point in time, t, a random error, e[i](t), is introduced into the ith record. Note that t now runs from 0 to R x T, rather than to T. The elements { e[i](t) } have zero mean and unknown variance.

3. The effect of an error introduced at time t extends, with reduced magnitude, for the next d time periods. (This is the assumption that motivated the creation of data set III, with d = 3.)

Under these assumptions the data point at time t, x[i](t), can be represented as a linear function that combines the systematic component, the error introduced at time t, and past error terms. Box and Jenkins refer to such models as mixed autoregressive and moving

average (ARMA) models. A slightly more general assumption is to assume that there is a slow drift in the regular process throughout the course of the experiment. This is called the seasonally adjusted ARMA model. The basic idea is to make an estimate of the observation at time t by considering the observaton at time t-1 <u>and</u> the observation at time t-T, i.e. the observation at the same point in the previous time cycle. The prediction equation used here was

(7) $x(t) = (x(t-1) - a\, e(t-1) + e(t)) + (x(t-T) - x(t-T-1))$

$+(b\, e(t-T) + a\, b\, e(t-T-1)) + w.$

where the subscript [i] has been dropped for convenience of notation. The first parenthesized term on the right hand side of (7) indicates that the observation at time t will include a carryforward term from time t-1. The second parenthesized term represents an estimated of the expected change from time t-1 to time t, based on observations at the corresponding point in the previous cycle. The third parenthesized term, and the "w" term, represent adjustments due to the fact that the observations for the previous time period will have contained error, and due to inaccuracy of prediction.

III, using the amplitudes with and without the general term, as described in discussing Table 2 above. The COFAM analysis reverses the PCA analysis in an interesting sense. Whereas PCA was most accurate in estimating the component amplitude including the general term, COFAM most accurately estimates the amplitudes without the general term.

TABLES 3 & 4

ANALYSIS OF ERPS USING TIME SERIES PROCEDURES

We next turn to problems in ERP analysis that are produced by violations of the assumption that the component wave forms are constant over records. This problem will be approached in a rather different way.

Conceptually, ERP analyses should be conducted on the records obtained from a single stimulus presentation. This is seldom practical due to the low signal:noise ratio in single trial records. In order to bring the signal:noise ratio to a manageable level, the unit of analysis is usually the record produced by averaging over several "equivalent" trials. Averaging over as many as 100 trials would not be an unusual

procedure. Providing that the records can be regarded as random samples from the same population, averaging can produce any desired improvement in the signal:noise ratio. The problem, though, *is* that it is extremely difficult to produce truly equivalent trials. Learning, fatigue, and habituation all can alter the forms of the component waves. To make matters worse, the various component forms may be differentially susceptible to alteration. For this reason, it is desirable to average over a small number of trials that are believed to have been recorded in, as nearly as possible, theoretically equivalent conditions. This goal is obviously at odds with the goal of achieving a high signal:noise ratio by averaging.

We explore here an alternative to the simple averaging method. The alternative, *Time Series Analysis* (TSA), was originally developed as a tool in econometrics (Box and Jenkins, 1976). A notational system somewhat different from the one used previously is required to explain the technique. The basic idea is to represent an ERP study <u>on a single case</u> as a sequence of T x R observations, where a stimulus is presented once every T time *periods*, for R repetitions. That is, the stimulus is presented at time 0, T, 2T, ... (R-1)T, and observations are made for the

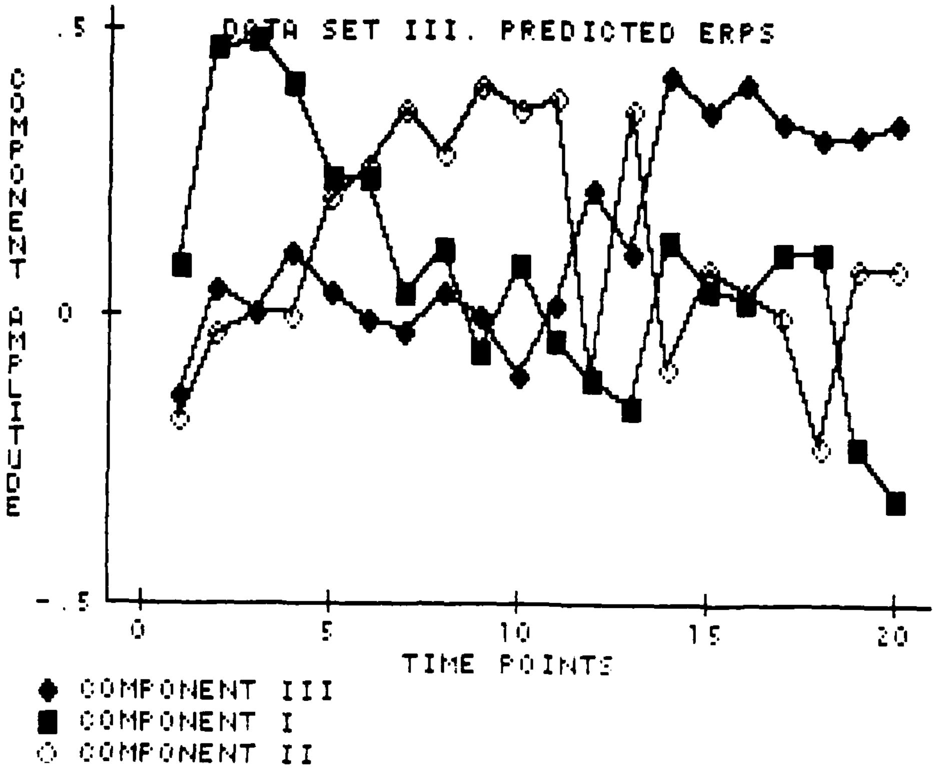

DATA SET III. PREDICTED ERPS
COMPONENT AMPLITUDE
.5
0
-.5
0
5
10
15
20
TIME POINTS
COMPONENT III
COMPONENT I
COMPONENT II

Figure 10B

END

FILMED

4-85

DTIC